

Classical and Quantum Tensor Product Expanders

M. B. Hastings

*Center for Nonlinear Studies and Theoretical Division,
Los Alamos National Laboratory, Los Alamos, NM, 87545*

A. W. Harrow

Department of Computer Science, University of Bristol, Bristol, U.K.

We introduce the concept of quantum tensor product expanders. These generalize the concept of quantum expanders, which are quantum maps that are efficient randomizers and use only a small number of Kraus operator. Quantum tensor product expanders act on several copies of a given system, where the Kraus operators are tensor products of the Kraus operator on a single system. We begin with the classical case, and show that a classical two-copy expander can be used to produce a quantum expander. We then discuss the quantum case and give applications to the Solovay-Kitaev problem. We give probabilistic constructions in both classical and quantum cases, giving tight bounds on the expectation value of the largest nontrivial eigenvalue in the quantum case.

I. BACKGROUND: CLASSICAL AND QUANTUM EXPANDERS

A. Definitions

The concept of t -designs[1] provides a way of randomizing quantum states. For example, a 1-design is a set of unitaries $\{U_k\}$, where $k = 1, \dots, K$, such that the average over the set takes any input state to a maximally mixed state. A 2-design is a set of unitaries such that applying $U_k \otimes U_k$ to a state on a bipartite system generates the twirling operation[2]. Quantum expanders, as studied in Hamiltonian complexity[3], computer science[4], and quantum information theory[5], provide a way of approximately realizing a 1-design by repeatedly applying a completely positive map built out of a small number of unitaries. In this paper, we introduce the concept of “tensor product expanders”, which generalize this result and give us a way to approximately realize t -designs. We also discuss the classical case, and show that classical tensor product expanders can be used to generate quantum expanders.

Quantum expanders are a quantum analogue of expander graphs[8]. In the quantum case, we consider a completely positive, trace preserving map

$$\mathcal{E}(M) = \sum_{s=1}^D A^\dagger(s) M A(s), \quad (1)$$

where the number of Kraus operators D is relatively small and the map \mathcal{E} has a spectral gap between the largest eigenvalue, equal to unity, and the next largest eigenvalue¹. We write the spectrum of \mathcal{E} as $\lambda_1, \lambda_2, \dots$ with $\lambda_1 = 1$ and λ_2, \dots all bounded in absolute value by some $\lambda < 1$. We can equivalently consider the operator $\hat{\mathcal{E}} := \sum_{s=1}^D A(s) \otimes A(s)^*$.

In this paper we consider the case in which the operators $A^\dagger(s)$ are proportional to unitary operators:

$$A(s) = \frac{1}{\sqrt{D}} U(s). \quad (2)$$

Then the expander map can be implemented by choosing s uniformly at random from $\{1, \dots, D\}$, and then applying $U(s)$ to the quantum state. The natural generalization of this process, in which we consider k copies of a quantum system, choose a unitary at random, and apply the unitary to all k copies, will be called a k -copy tensor product expander. We will show that these give a way to approximate t -designs for $t = k$.

Random walks on expander graphs can be viewed similarly, as acting on a distribution with a randomly chosen permutation matrix. Consider a directed graph, where each node has D edges leaving it. Label the edges from 1 up to D such that each label appears exactly once among the incoming edges of each vertex and exactly once among the outgoing edges of each vertex. Then, for each edge label s , $1 \leq s \leq D$, define a permutation π_s , where $\pi_s(i) = j$ if

¹ In the non-Hermitian case discussed below, we define the gap instead to be one minus the second-largest singular value of the map \mathcal{E} .

a directed edge with label s goes from node i to node j . Then, given a random walk on the graph, the probability distribution $p(i)$ changes in a single step by

$$p(i) \rightarrow \frac{1}{D} \sum_{s=1}^D \sum_{j=1}^N P(s)_{ij} p(j), \quad (3)$$

where $P(s)$ is the permutation matrix corresponding to the permutation π_s ; i.e. $P(s)_{ij} = 1$ if $\pi_s(j) = i$ and 0 otherwise.

Hermitian expanders: It is sometimes convenient to guarantee that an expander we construct is Hermitian. To obtain Hermitian \mathcal{E} in the quantum case, we impose

$$U(s + D/2) = U(s)^\dagger. \quad (4)$$

Similarly, in the classical case, we impose

$$\pi_s = \pi_{s+D/2}^{-1} \quad (5)$$

This turns the directed graph into an undirected graph. For notational convenience, we identify $s + D$ with s throughout this paper, so that s is a periodic variable with period D . Note that this constraint (4) requires that D be even. There do exist other ways to construct Hermitian expanders with odd D , if for some s we have $U(s) = U(s)^\dagger$.

B. Application to state randomization

For classical expanders, an important implication of the spectral gap is that random walks on an expander graph rapidly approach the stationary distribution. Similarly, quantum expanders can be shown to be rapid mixing. This has application to the problem of *state randomization*, in which classical randomness is used to map a quantum state to an output that is close in trace distance to the maximally mixed state. Ideally the constructions would be [computationally] efficient, meaning they run in time polynomial in the number of qubits, and would use as few random bits as possible.

To make this concrete, suppose that \mathcal{E} is Hermitian and unital with gap $1 - \lambda$, and consider a quantum state ρ . We wish to bound the trace norm distance between the maximally mixed state and the state $\mathcal{E}^m(\rho)$ obtained by acting on ρ with some high power, m , of the map \mathcal{E} . The calculation exactly follows the classical case. We begin by bounding the ℓ_2 distance. For a matrix A , define $\|A\|_2 = \sqrt{\text{tr } A^\dagger A}$ and $\|A\|_1 = \text{tr } |A| = \text{tr } \sqrt{A^\dagger A}$. Then

$$\left\| \mathcal{E}^m(\rho) - \frac{\mathbb{1}}{N} \right\|_2^2 \leq |\lambda|^{2m}, \quad (6)$$

as may be shown by writing ρ as a linear combination of eigenvectors of \mathcal{E} , and then by Cauchy-Schwartz,

$$\left\| \mathcal{E}^m(\rho) - \frac{\mathbb{1}}{N} \right\|_1 \leq \sqrt{N} |\lambda|^m. \quad (7)$$

Thus, to obtain a given bound on the trace norm distance ϵ , it suffices to take

$$m \geq \log_\lambda(\epsilon/\sqrt{N}). \quad (8)$$

This implies that the set of unitaries, consisting of all unitaries of the form $U(s_1)U(s_2)\cdots U(s_m)$, gives an ϵ -approximate 1-design using

$$K := D^m = \left(\frac{N}{\epsilon^2} \right)^{\frac{1}{2} \log_{1/\lambda}(D)} \quad (9)$$

unitaries.

The exponent $\frac{1}{2} \log(D)/\log(1/\lambda)$ can be thought of as a measure of the efficiency of an expander, meaning the number of bits of randomness it requires to achieve a certain amount of state randomization. Before showing how to evaluate $\frac{1}{2} \log(D)/\log(1/\lambda)$, we review other methods of ℓ_1 state randomization. The simplest is to apply one of N^2 generalized Pauli operators. This can be done efficiently (i.e. in time $\text{poly} \log(N)$) and perfectly randomizes any state (i.e. $\epsilon = 0$). However, it uses far more randomness than necessary when $\epsilon > 0$. Choosing $K = O(N\epsilon^{-2} \log(1/\epsilon))$

random unitaries was shown to suffice in [10], improving a result of [11] (both of which in fact addressed the more difficult problem of ℓ_∞ state randomization). Similarly an efficient $K = 4N\epsilon^{-2}$ construction was given in [12], which uses less randomness than the efficient constructions of [13] and even than the inefficient constructions based on random unitaries. We note in passing that the constructions in [12, 13] are based on expanders with $\lambda = \epsilon/\sqrt{N}$ and $D = K$.

An expander-based state randomization scheme will be efficient if the underlying expander is efficient and the number of unitaries it uses will be given by (9). Unfortunately $\frac{1}{2}\log(D)/\log(1/\lambda)$ is larger than 2 for all known efficient constant-degree expander constructions[5, 6, 7] (e.g. for the Margulis expander[6], it is ≈ 8.4 , and for the zig-zag product[5] it is $2 + o(1)$). However, if $U(1), \dots, U(D/2)$ are chosen at random with $U(s + D/2) = U(s)^\dagger$ then Ref. [19] showed that with high probability $\frac{1}{2}\log(D)/\log(1/\lambda) \approx 1 + \mathcal{O}(\log(N)N^{-1/6}) + 2/\log(D)$, and thus that K is within a small multiplicative factor of N/ϵ^2 .

We summarize the above discussion as follows:

Theorem 1 *For any N and any $\epsilon > 0$, consider a set of unitaries $U_1, \dots, U_K \in \mathcal{U}_N$, which are taken to be strings of unitaries drawn from a set of $D/2$ unitaries $U(1), \dots, U(D/2)$ and their conjugates for any $D \geq 4$. Then for most choices of $U(1), \dots, U(D/2)$, choose the string length such that*

$$K = \left(\frac{N}{\epsilon^2}\right)^{1+O(N^{-1/6}\log(N))+2/\log(D)} \quad (10)$$

and

$$\left\| \frac{1}{K} \sum_{s=1}^K U_s \rho U_s^\dagger - \frac{\mathbb{I}}{N} \right\|_1 \leq \epsilon,$$

for all N -dimensional density matrices ρ .

If we take $D \approx 4N/\epsilon^2$ then Theorem 1 can be thought of as tightening the analysis of random unitaries from [10, 11, 12], so that only $(4 + o(1))N/\epsilon^2$ random unitaries are necessary. This shows that Haar-uniform unitaries require almost exactly the same amount of randomness as the construction of [12], although they have the substantial disadvantage of requiring $\text{poly}(N)$ time to implement instead of $\text{poly}(\log(N))$ time. Since $\lambda \geq (2\sqrt{D-1}/D - O(1/N)) \cdot (1 - O(\log \log(N)/\log(N)))$ for any quantum expander that includes its own inverses [19], one can show that $4N/\epsilon^2$ is the minimum possible values of K for any expander-based randomizing map.

Apart from random unitaries and the large- D constructions of [12, 13], we know of one other class of quantum expanders for which $\frac{1}{2}\log(D)/\log(1/\lambda) \approx 1$. These are obtained by applying the prescription of [7] to the $SU(2)$ expanders described by Lubotsky, Phillips and Sarnak in [14]. Such expanders exist for any N whenever D is odd and $2D - 1$ is prime, and satisfy $\lambda = 2\sqrt{D-1}$ exactly. Thus, they provide another $K \approx 4N/\epsilon^2$ method of performing state randomization. However, the only claimed efficient construction of these expanders[15] has an incomplete proof.

In the non-Hermitian case, (6) holds when λ is the second-largest singular value of an expander. If $U(1), \dots, U(D)$ are chosen uniformly at random, then [19] proved that with high probability the singular values of \mathcal{E}^m for $m = \mathcal{O}(N^{1/6})$ are bounded by $N^2(1/\sqrt{D})^m(1 + o(1))$. This implies that the second-largest *eigenvalue* of \mathcal{E} is $\leq \frac{1}{\sqrt{D}}(1 + O(\log(N)N^{-1/6}))$, but does not yield meaningful bounds on the second-largest singular value of \mathcal{E} . Indeed, Tobias Osborne has pointed out that when $m = 1$ and $D = 2$, the second largest singular value is equal to unity. If \mathcal{E}^m turned out to have singular values nearly equal to $D^{-m/2}$ then it would imply that $\approx N/\epsilon^2$ random unitaries sufficed to ϵ -randomize a state.

We now turn to tensor product expanders, considering classical tensor product expanders in Section II and quantum tensor product expanders in Section III. The mixing analysis above generalizes in the tensor product case to give approximate t -designs. We will describe randomized constructions of both classical and quantum tensor product expanders. Our basic tool to prove that a random construction gives an expander with high probability is the trace method (see, for example [8, 18]). The basic idea of the trace method is to bound eigenvalues of some linear operator by bounding the trace of high powers of that operator. For example, for a positive definite Hermitian operator whose two largest eigenvalues are equal to unity and to λ , the trace of the m^{th} power is at least equal to $1 + \lambda^m$, so by bounding the trace we bound λ . We focus on high powers of the operator so that the trace will be dominated by the largest eigenvalues. The trace method will be adapted, with slight modifications, to the various cases, depending on whether classical or classical and quantum, and depending on whether we consider an expander and or a tensor product expander.

II. CLASSICAL TENSOR PRODUCT EXPANDERS

In this section we define classical tensor product expanders, and give a random construction of them. We then show an application of them to constructing quantum expanders.

A. Preliminaries, Definitions and Applications

We define an (N, D, λ, k) classical k -copy tensor product expander to be a set of N -by- N permutation matrices $P(s)$, $1 \leq s \leq D$, with the property that the matrix L , defined by

$$L_k = \frac{1}{D} \sum_{s=1}^D P(s)^{\otimes k} \quad (11)$$

has some number, f_k^N , eigenvalues equal to unity, with f_k^N defined below, and then all other eigenvalues less than or equal to λ in absolute value. (Again, if L_k is non-Hermitian then we consider its singular values.)

We can obtain Hermitian operators L_k by considering D even, and imposing $P(s + D/2) = P(s)^\dagger$. To obtain Hermitian L_k for D odd, we can instead impose $P(s) = P(s)^\dagger$; that is, the permutation matrices correspond to perfect matchings. Both models corresponds to models of random graphs for $k = 1$ discussed in [9].

These expanders can also be defined by graphs with N^k nodes, labelled (n_1, n_2, \dots, n_k) , where $1 \leq n_i \leq N$. There is an edge from one node (n_1, \dots, n_k) to another node (n'_1, \dots, n'_k) if and only if one of the given permutations sends $n_1 \rightarrow n'_1, \dots, n_k \rightarrow n'_k$. We refer to this graph as G_k . Alternatively, we can regard n_1, \dots, n_k as k different random walkers executing a correlated random walk on the original graph.

The function f_k^N is defined to be equal to the number of unit eigenvalues of the operator

$$\frac{1}{N!} \sum_{\pi \in S_N} P_\pi^{\otimes k} \quad (12)$$

where the sum ranges over *all* permutations π , and P_π is the permutation matrix corresponding to permutation π . Since this operator performs an average over a group action, it is a projector. Applying it to a computational basis state $|n_1, \dots, n_k\rangle$ maps it to the superposition of all $|n'_1, \dots, n'_k\rangle$ such that $n'_i = n'_j$ iff $n_i = n_j$. Thus we can represent eigenstates by partitions of $\{1, \dots, k\}$ into $\leq N$ blocks, such that indices are equal within blocks and unequal across blocks. For example, $f_1^N = 1$, $f_2^N = 2$ (corresponding to the sum of all states with $n_1 = n_2$ and the sum of all states with $n_1 \neq n_2$), $f_3^N = 5$ (corresponding to the possibilities $n_1 = n_2 = n_3$, $n_1 = n_2 \neq n_3$, $n_1 = n_3 \neq n_2$, $n_2 = n_3 \neq n_1$, and $n_1 \neq n_2 \neq n_3 \neq n_1$), and so on. Note that if $N \geq k$ then the constraint that there be $\leq N$ blocks becomes superfluous, and f_k^N becomes simply the k^{th} Bell number B_k , which counts the total number of ways of partitioning a k -element set.

Any matrix L_k of the form (11) is block diagonal with f_k^N different blocks depending on the symmetry of the elements n_1, \dots, n_k under permutation; we call these subspaces $S_1, S_2, \dots, S_{f_k^N}$. By the arguments of the above paragraph, we can write the projector in (12) as

$$\sum_{a=1}^{f_k^N} |u_a\rangle \langle u_a|,$$

for some unit vectors $|u_a\rangle \in S_a$. These $|u_a\rangle$ are unit eigenvalues not only of (12) but also any L_k .

Rapid mixing: Given the spectral gap, repeatedly applying a classical tensor product expander many times (of order $k \log(N)$) generates an approximately k -wise independent permutation. This means that the results of applying it to k distinct elements are almost indistinguishable from applying a single permutation to each of the k elements. More precisely, given an initial probability distribution, p , in any of the f_k^N different subspaces S_a , we have

$$\|L_k^m p - u_a\|_1 \leq \sqrt{N^k} |\lambda|^m, \quad (13)$$

where u_a is the l_1 normalized eigenvector with eigenvalue unity in this subspace. This approach towards generating k -wise independent permutations has also been considered in [16].

Expanders are not always tensor product expanders. The requirement that a set of permutations form a tensor product expander for $k > 1$ copies is more stringent than the requirement for $k = 1$ copy, as it implies that the correlations between elements are destroyed by the expander. For an example of a classical expander that does

not give a tensor product expander, consider any set of D permutation matrices, $P(s)$, on N elements that gives a classical expander. Define a new set of permutation matrices, $P'(s)$, on $2N$ elements, such that $P'(s) = P(s) \oplus P(s)$ for $s = 1, \dots, D$. Finally, define the permutation $P'(D+1)$ which sends i to $i+N$ if $i \leq N$, and sends i to $i-N$ if $i > N$. Then, these $D+1$ different permutation matrices define a $k=1$ expander (they simply correspond to two copies of the original graph, with the possibility of moving between the two copies by using permutation matrix $P'(D+1)$), but does not define a $k=2$ expander: if two walkers, n_1, n_2 originally are in the same copy as each other, then they remain in the same copy.

Another example comes from Cayley graphs. If G is a group with generators g_1, \dots, g_D then the Cayley graph on G is defined by taking $N = |G|$ and $P(s)|g\rangle = |g_s g\rangle$ for $s = 1, \dots, D$. There are many Cayley graph expanders known (c.f. Section 11 of [8]), but applying $P(s) \otimes P(s)$ to any $|g\rangle \otimes |h\rangle$ produces a new state $|\tilde{g}\rangle \otimes |\tilde{h}\rangle$ with $\tilde{g}^{-1}\tilde{h} = g^{-1}h$. Thus, no Cayley graph expander can be a tensor product expander unless it is modified in some way.

The limit of large k : Observe that any k -copy tensor product expander is also a k' -copy tensor product expander for all $k' \leq k$. On the other hand, even if $k > N$ then the k walkers can still occupy only at most N positions. Thus if a map is an N -copy tensor product expander then it is also a k -copy tensor product expander for all k .

An equivalent condition to $\{\pi_1, \dots, \pi_D\} \subset \mathcal{S}_N$ being an N -tensor product expander is that the Cayley graph generated by $\{\pi_1, \dots, \pi_D\}$ is an expander. The spectrum of this Cayley graph is identical (up to multiplicity) to that of L_k for all $k \geq N$ (with $P(s)$ defined to be P_{π_s}).

B. Random permutations are tensor product expanders

The question then naturally arises whether $k > 1$ tensor product expanders actually exist. Of course there is a trivial $D = N!$ construction where we take $\{\pi_1, \dots, \pi_N\} = \mathcal{S}_N$ and achieve $\lambda = 0$ for all k . We would prefer, though, that $D = O(1)$. The construction of [16] nearly achieves this with $D = \text{poly}(\log(N))$ and $\lambda = 1 - 1/\text{poly}(k, \log(N))$. For a constant degree construction, we can use Kassabov's expander[17] on \mathcal{S}_N . This achieves $D = O(1)$ and λ equal to a constant strictly smaller than 1 for all N and k . Additionally, it can be implemented in time $\text{poly}(\log(N))$.

In this section, we give a randomized construction of tensor product expanders for any even $D \geq 4$ and with $\lambda \approx \lambda_H^{\frac{1}{k+1}}$, where

$$\lambda_H := \frac{2\sqrt{D-1}}{D}. \quad (14)$$

Theorem 2 Choose $\pi_1, \dots, \pi_{D/2} \in \mathcal{S}_N$ at random and then take $\pi_{s+D/2} = \pi_s^{-1}$. Let $P(s) = P_{\pi_s}$. For any k , let λ denote the $f_k^N + 1^{\text{st}}$ largest eigenvalue of L_k . Then for any $c > 1$,

$$\Pr\left[\lambda \geq c\left(\lambda_H^{\frac{1}{k+1}} + O\left(\frac{\log(k) + \log(\log(N))}{\log(N)}\right)\right)\right] \leq c^{(-k+1)\log_{1/\lambda_H}(N)}, \quad (15)$$

where $\Pr[\dots]$ denotes probability and λ_H depends on D as given in Eq. (14).

Note that since $\lambda_H^{1/(k+1)}$ converges to unity as k becomes large, the result (15) is only meaningful for $k = \mathcal{O}(\log(N)/\log(\log(N)))$. Constants depending on D are also hidden inside of the $O()$ notation. The result is likely far from optimal, since numerical studies indicate that for fixed k and large N , the largest non-trivial eigenvalue λ approaches λ_H . This result for the case $k=1$ was only recently proven[9]. Our proof, which gives a weaker bound on the expectation value of λ roughly follows the presentation of the trace method in [8, 18], with some modifications.

Proof of Theorem 2: We will apply the trace method separately in each of the subspaces S_a . It suffices to consider only one such subspace S_a , the subspace $S_{f_k^N}$ in which all of the n_1, n_2, \dots, n_k differ from each other, since every eigenvalue of L_k is an eigenvalue of L_k restricted to $S_{f_k^N}$. For example, consider the case $k=2$. We have two different subspaces, one with $n_1 = n_2$ and one with $n_1 \neq n_2$. The eigenvectors of the first subspace, of the form $\sum_i p(i)|i\rangle|i\rangle$, correspond to eigenvectors of L_1 of the form $\sum_i p(i)|i\rangle$. Given such an eigenvector, we can construct an eigenvector in the second subspace equal to $\sum_i \sum_{j \neq i} p(i)|i\rangle|j\rangle$ with the same eigenvalue, as claimed.

Let $\mathbb{E}[\dots]$ denote an average over different choices of permutation matrices. Then for any even m ,

$$\mathbb{E}[|\lambda|] \leq (\mathbb{E}[\text{tr}(L_k^m R)] - 1)^{1/m}, \quad (16)$$

where R is the projector onto the given subspace. The expectation value $\mathbb{E}[\text{tr}(L_k^m R)]$ equals

$$\left(\frac{1}{D}\right)^m \sum_{s_1=1}^D \sum_{s_2=1}^D \dots \sum_{s_m=1}^D \mathbb{E}[\text{tr}(P(s_1)P(s_2)\dots P(s_m)R)]. \quad (17)$$

If for some i we have $s_i = s_{i+1} + D/2$, then $P(s_i)P(s_{i+1}) = \mathbb{1}$, and we can remove that pair of permutation matrices from the trace above. Similarly, if $s_m = s_1 + D/2$, then we can remove the first and last permutation matrices from the trace, exploiting the cyclic invariance of the trace and the vanishing commutator $[P(s), R] = 0$. We can consider these operations as acting on a word s_1, s_2, \dots, s_m on an alphabet $\{1, \dots, D\}$. We define a reduced word by removing pairs of letters of the form $s, s + D/2$. Similarly, if the word ends with a letter s and begins with a letter $s + D/2$, we remove this pair also. We repeat these removals until no further removals are possible. The result is a reduced word of length $m^0 \leq m$; the resulting sequence we write $s'_1, s'_2, \dots, s'_{m^0}$. There are at most

$$(D-1)^{m/2} 2^m = D^m \lambda_H^m \quad (18)$$

choices of s_1, \dots, s_m which give $m^0 = 0$; the number of these choices is equal to D^m times the return probability of a random walk of length m on a Cayley tree of degree D . For these choices, we have $\mathbb{E}[\text{tr}(P(s_1)P(s_2)\dots P(s_m)R)] = \text{tr}(R) \leq N^k$.

We now consider the other choices of s_1, \dots, s_m , where $m^0 > 0$. In general,

$$\mathbb{E}[\text{tr}(P(s'_1)P(s'_2)\dots P(s'_{m^0})R)] \leq N^k \mathbb{E}[\text{tr}(P(s'_1)P(s'_2)\dots P(s'_{m^0})R_{1,2,\dots,k})], \quad (19)$$

where $R_{1,2,\dots,k}$ projects onto the state with $n_1 = 1, n_2 = 2, \dots, n_k = k$. To compute this expectation value, we define $v_0^a = a$, for $1 \leq a \leq k$. Then, define v_i^a , for $i \geq 1$ and $1 \leq a \leq k$, to be $\pi_{s'_i}(v_{i-1}^a)$. Then, the probability that $v_{m^0}^a = a$ for all a is equal to the desired result. We compute this probability as follows. Consider this as happening sequentially, where first we define v_1^a for all a , then we define v_2^a , and so on. We say that a choice of v_i^a is “free” if at no previous step $j < i$ did we compute $\pi_{s'_j}(v_{j-1}^a)$ with $s'_j = s'_i$ and $v_{j-1}^a = v_{i-1}^a$. If a choice of v_i^a is free, and if t values of $\pi_{s'_i}$ have been previously revealed, then we can simply pick v_i^a at random from the $N - t$ possibilities, thus revealing some of the information about the permutation $\pi_{s'_i}$, and increasing t by one for that permutation. If a choice is not free, then it is “forced”, in which case we have no choice about the value of $\pi_{s'_i}(v_{i-1}^a)$.

We say that a coincidence occurs at step i for walker a if this is a free step and the randomly selected vertex coincides with a previously selected vertex (previously selected by *any* of the walkers). Note that for $v_{m^0}^a$ to equal a for all a , we must have at least k coincidences. There are two cases: either there are at least $k + 1$ coincidences, or else there are exactly k coincidences.

The probability of there being at least $k + 1$ coincidences can be computed as follows. Let i_1, i_2, \dots, i_{k+1} be the steps of the first $k + 1$ coincidences and a_1, a_2, \dots, a_{k+1} be the corresponding walkers. The probability of having these coincidences for given i_1, \dots and a_1, \dots is bounded by $(mk/(N - mk))^{k+1}$. Summing over all possible steps and walkers, we find that the probability of having at least $k + 1$ coincidences is bounded by

$$m^{k+1} k^{k+1} (mk/(N - mk))^{k+1}. \quad (20)$$

If there are exactly k coincidences, then each walker has exactly one coincidence given that $v_{m^0}^a = a$ for all a . There are two possibilities: either all of the coincidences occur on the last step, or at least one coincidence does not occur on the last step. The probability of the first case is at most $(1/(N - mk))^k$. If at least one coincidence does not occur on the last step, then let walker b be the first walker to have a coincidence, occurring on step j . Note that each of the vertices $1, \dots, a$ must be the randomly selected vertex on exactly one coincidence, again given that $v_{m^0}^a = a$ for all a . Because there are no further coincidences for walker b , we have $s'_i = s'_{i+j}$ for all i . The fraction of reduced words of length m_0 that obey this constraint for given $j \leq m_0/2$ is at most $(D - 1)^{-m_0/2}$. The fraction of words that have a reduced word of length m_0 is at most $(D - 1)^{m_0/2} \lambda_H^m$. Therefore, the fraction of words that have a reduced word obeying this constraint, after summing over j , is at most $m \lambda_H^m$. The probability of having these coincidences is bounded by $(m/(N - mk))^k$, where the factor of m arises from the choice of step on which the coincidence occurs (this is in fact a large overestimate). The product of these probabilities is $m \lambda_H^m (m/(N - mk))^k$. The total of these two possibilities is

$$(1/(N - mk))^k + (m/(N - mk))^k m \lambda_H^m. \quad (21)$$

Adding the sum of the expectation value over words with $m^0 = 0$ (which is bounded by $N^k \lambda_H^m$ by Eq. (18) to N^k times the sum of (20,21), we find that

$$\mathbb{E}[\text{tr}(P(s'_1)P(s'_2)\dots P(s'_{m^0})R)] \leq N^k \lambda_H^m + N^k m^{k+1} k^{k+1} (mk/(N - mk))^{k+1} + (N/(N - mk))^k + (Nm/(N - mk))^k m \lambda_H^m. \quad (22)$$

and therefore

$$\begin{aligned} & \mathbb{E}[\text{tr}(P(s'_1)P(s'_2)\dots P(s'_{m^0})R)] - 1 \\ & \leq N^k \lambda_H^m + N^k m^{k+1} k^{k+1} (mk/(N - mk))^{k+1} + [(N/(N - mk))^k - 1] + (Nm/(N - mk))^k m \lambda_H^m \\ & = N^k \lambda_H^m + N^k m^{k+1} k^{k+1} (mk/(N - mk))^{k+1} + \mathcal{O}(mk^2/N) + (Nm/(N - mk))^k m \lambda_H^m. \end{aligned} \quad (23)$$

We pick

$$m = (k + 1) \log_{1/\lambda_H}(N) \quad (24)$$

to minimize this expectation value, finding

$$(\mathbb{E}[\text{tr}(P(s'_1)P(s'_2)\dots P(s'_{m^0})R)] - 1)^{1/m} \leq \lambda_H^{1/(k+1)} (\mathcal{O}(mk))^{(k+1)/m}. \quad (25)$$

Applying Markov's inequality then yields the proof of the Theorem. \blacksquare

C. Quantum expanders from classical tensor product expanders

One application of $k = 2$ classical tensor product expanders is to constructing quantum expanders. We give two constructions.

The first approach was introduced, but not formally analyzed, in [3]. Let $P(s)$ be a set of random permutation matrices defining a $k = 2$ tensor product expander, as in the random construction of a $k = 2$ tensor product expander above. Then, define $\sigma(s)$, for $s = 1 \dots D$, to be a diagonal matrix. For $s = 1, \dots, D/2$ we choose $\sigma(s)$ to have diagonal entries ± 1 chosen independently at random and we choose $\sigma(s + D/2) = P(s)\sigma(s)P(s)^\dagger$. Then, in [3] it was shown numerically that the A matrices,

$$A(s) = \frac{1}{\sqrt{D}} P(s) \sigma(s), \quad (26)$$

define a quantum expander with high probability. Note that the choice of $\sigma(s + D/2)$ is such that $A(s + D/2) = A(s)^\dagger = (1/\sqrt{D})\sigma(s)P(s)^\dagger$ so that this is a Hermitian expander because $P(s) = P(s + D/2)^\dagger$. Numerically, λ was observed to approach λ_H for large N . We now prove that we do indeed get a quantum expander with high probability, but with a weaker bound on λ_H .

Theorem 3 *Choose $\pi_1, \dots, \pi_{D/2} \in \mathcal{S}_N$ at random and then take $\pi_{s+D/2} = \pi_s^{-1}$. Let $P(s) = P_{\pi_s}$. Choose $\sigma(s)$ as described above. Let λ denote the second largest eigenvalue of the map with Kraus operators given by the matrices $A(s)$ in Eq. (26). Then, for any $c > 1$,*

$$\Pr\left[\lambda \geq c\left(\lambda_H^{\frac{1}{3}} + O\left(\frac{\log(\log(N))}{\log(N)}\right)\right)\right] \leq c^{-3 \log_{1/\lambda_H}(N)}. \quad (27)$$

The Hermitian, completely positive map \mathcal{E} defined by the A matrices in (26) sends a diagonal matrix to a diagonal matrix and an off-diagonal matrix to an off-diagonal matrix. So, we consider the spectrum of \mathcal{E} in the diagonal and off-diagonal sectors separately. In the diagonal sector, the spectrum of \mathcal{E} is the same as that of the $k = 1$ expander defined by the given permutation matrices, and hence has a gap between the largest eigenvalue, equal to unity, and the next largest eigenvalue.

The off-diagonal sector requires a little more work. We again use the trace method. Let λ be the largest eigenvalue in absolute value in the off-diagonal sector. Let $M(i, j)$ be an N -by- N dimensional matrix with a one in the i^{th} row and j^{th} column, and zeroes everywhere else, so that these form a basis for the space of N -by- N matrices. The $M(i, j)$ with $i \neq j$ form a basis for the space of off-diagonal matrices. Define (M, N) to be an inner product on the space of N^k -by- N^k dimensional matrices by $(M, N) = \text{tr}(M^\dagger N)$. Then for any even m ,

$$\mathbb{E}[|\lambda|] \leq \left(\mathbb{E}\left[\sum_{i \neq j} \left(M(i, j), \mathcal{E}^m(M(i, j))\right)\right]\right)^{1/m}. \quad (28)$$

Note that compared to Eq. (16), a factor of unity is not subtracted from the expectation value on the right-hand side of Eq. (28).

The evaluation of the right-hand side of Eq. (28) proceeds analogously to that of Eq. (16). The computation in the case $m^0 = 0$ is identical. In the case $m^0 > 0$, we again define coincidences and paths. The only difference is that now rather than just computing the probability that $v_{m^0}^a = a$ for all $a = 1, 2$, the paths come in with signs which may be plus or minus one. This can only reduce the contribution of the terms with $m^0 > 0$. We bound the case with $k + 1$ coincidences as before. We also bound the case with k coincidences not all occurring on the last step as before. The only difference is the case in which all coincidences happen on the last step $i = m^0$. The probability of this happening

is $(1/N)^2$. The sign, however, is completely random; it is equally likely to be plus or minus one. Thus, the paths with exactly k coincidences, all occurring on step $i = m^0$, contribute zero to the expectation value (28). Thus,

$$\mathbb{E}[|\lambda|^m] \leq (N^k + m)\lambda_H^m + N^k m^{k+1} k^{k+1} (mk/(N - mk))^{k+1}. \quad (29)$$

Picking m as before, we find that $\mathbb{E}[|\lambda|] \leq \lambda_H^{1/3} (1 + \mathcal{O}(\log(\log(N))/\log(N)))$. Applying Markov's inequality yields the theorem. ■

We now describe our second construction of a quantum expander from a classical tensor product expander.

Theorem 4 *Suppose $\{P(1), \dots, P(D)\}$ form a $(N, D, 1 - \epsilon, 2)$ classical tensor product expander (i.e. $k = 2$). Assume that $N \geq 2$. Let*

$$Z = \sum_{j=1}^N |j\rangle\langle j| e^{\frac{2\pi i j}{N}}$$

and $p = 1/(1 + \epsilon)$. Define a quantum operation $\mathcal{E}(M)$ with $D + 1$ Kraus operators $\sqrt{\frac{p}{D}}P(1), \dots, \sqrt{\frac{p}{D}}P(D), \sqrt{1 - p}Z$. Then \mathcal{E} is a $(N, D + 1, 1 - \frac{\epsilon}{48})$ quantum expander.

Thus, any constant-gap classical 2-TPE can be used to construct a constant-gap quantum expander. No attempt has been made to optimize the constant 48, which we believe can be made arbitrarily close to one when N is large and ϵ is close to 1.

Note that $\sqrt{\frac{p}{D}}P(1), \dots, \sqrt{\frac{p}{D}}P(D), \sqrt{1 - p}Z$ is not in general Hermitian, but if $\{P(1), \dots, P(D)\}$ is Hermitian then $\{\sqrt{\frac{p}{D}}P(1), \dots, \sqrt{\frac{p}{D}}P(D), \sqrt{\frac{1-p}{2}}Z, \sqrt{\frac{1-p}{2}}Z^\dagger\}$ is a Hermitian $(N, D + 2, 1 - \epsilon/48)$ expander; this is proved by using the triangle inequality to relate its gap to the gap of the expander in Theorem 4.

Proof of Theorem 4: The idea is that the classical TPE randomizes the diagonal elements of the density matrix simply because it is an expander, and it randomizes the off-diagonal elements because it is a $k = 2$ TPE. Next the phase operation Z adds a phase to the off-diagonal elements so that they are no longer fixed by the classical TPE. Thus the only fixed state will be the identity matrix.

More formally, let $|\varphi_1\rangle = \frac{1}{\sqrt{N}} \sum_{i=1}^N |i\rangle|i\rangle$ and $|\varphi_2\rangle = \frac{1}{\sqrt{N(N-1)}} \sum_{i \neq j} |i\rangle|j\rangle$. These two states form an orthonormal basis for the invariant subspace of $\frac{1}{D} \sum_{s=1}^D P(s) \otimes P(s)$. Thus the fact that $P(1), \dots, P(D)$ form a 2-TPE implies the bound

$$\left\| \frac{1}{D} \sum_{s=1}^D P(s) \otimes P(s) - \varphi_1 - \varphi_2 \right\| \leq \lambda.$$

Next, a short calculation shows that $\langle \varphi_2 | Z | \varphi_2 \rangle = -1/(N - 1)$. Now apply the following Lemma to the subspace orthogonal to $|\varphi_1\rangle$.

Lemma 1 *Let Π be a projector and let X and Y be operators such that $\|X\| \leq 1$, $\|Y\| \leq 1$, $\Pi X = X \Pi = \Pi$, $\|(I - \Pi)X(I - \Pi)\| \leq 1 - \epsilon_X$ and $\|\Pi Y \Pi\| \leq 1 - \epsilon_Y$. Assume $0 < \epsilon_X, \epsilon_Y < 1$. Then for any $0 < p < 1$, $\|pX + (1 - p)Y\| < 1$. Specifically,*

$$\|pX + (1 - p)Y\| \leq 1 - \frac{\epsilon_Y}{12} \min(p\epsilon_X, 1 - p). \quad (30)$$

Setting $p = 1/(1 + \epsilon_X)$, we obtain

$$\|pX + (1 - p)Y\| \leq 1 - \frac{\epsilon_X \epsilon_Y}{12(1 + \epsilon_X)} \leq 1 - \frac{\epsilon_X \epsilon_Y}{24}. \quad (31)$$

The Lemma is proved in Appendix A. We apply the Lemma by taking $X = \frac{1}{D} \sum_{s=1}^D P(s) \otimes P(s) - \varphi_1$, $Y = Z \otimes Z^* - \varphi_1$ and $\Pi = \varphi_2$. Then plugging $\epsilon_X = \epsilon$ and $\epsilon_Y = 1 - 1/(N - 1) \geq 1/2$ into (31) completes the proof of Theorem 4. ■

III. QUANTUM TENSOR PRODUCT EXPANDERS

In this section we define quantum tensor product expanders and show that random unitaries provide a way of constructing tensor product expanders. We begin with some preliminaries and definitions, present applications to the Solovay-Kitaev problem of approximating unitaries by a string of elementary operations, and finally prove that random unitaries give tensor product expanders. The proof of this last statement begins in subsection III C; it closely follows [19] and should be read in conjunction with that paper.

A. Preliminaries, Definitions, and Applications

Suppose we have a collection of unitaries $\{U(1), \dots, U(D)\} \in \mathcal{U}_N$. Define a quantum operation \mathcal{E}_k that applies $U_s^{\otimes k}$ for $s \in \{1, \dots, D\}$ chosen uniformly at random. In other words

$$\mathcal{E}_k(M) = \frac{1}{D} \sum_{s=1}^D U(s)^{\otimes k} M (U(s)^\dagger)^{\otimes k}, \quad (32)$$

where M is an $N^k \times N^k$ matrix. Since an $N^k \times N^k$ matrix can also be viewed as an N^{2k} -dimensional vector, we can also interpret \mathcal{E}_k as a linear operator on an N^{2k} -dimensional vector space. Define this operator to be

$$\hat{\mathcal{E}}_k := \frac{1}{D} \sum_{s=1}^D U(s)^{\otimes k} \otimes (U(s)^*)^{\otimes k}. \quad (33)$$

Note that \mathcal{E}_k and $\hat{\mathcal{E}}_k$ are isospectral.

In previous work[4, 5, 19] \mathcal{E}_1 was said to be a (N, D, λ) quantum expander if the second-largest eigenvalue of $\hat{\mathcal{E}}_2$ was $\leq \lambda$. In fact, the definition of quantum expanders included even quantum operations that were not mixtures of unitaries, as long as they could be expressed using $\leq D$ Kraus operators. Here we will change notation from [4, 5, 19] slightly. We say that a set of unitaries $\{U(1), \dots, U(D)\}$ is a (N, D, λ, k) tensor product expander if the operator \mathcal{E}_k has F_k^N (defined below) eigenvalues equal to one, and all of its other eigenvalues have absolute value $\leq \lambda$. This differs from the notation of [4, 5, 19] in that the set of unitaries, rather than the quantum operation, constitutes the quantum expander². When N and D are understood, we sometimes simply say that $\{U(1), \dots, U(D)\}$ are a k -tensor product expander with gap $1 - \lambda$.

We define F_k^N to be the rank of the projector

$$\hat{T}_k := \int_{U \in \mathcal{U}_N} U^{\otimes k} \otimes (U^*)^{\otimes k} dU$$

or equivalently of the operation \mathcal{T}_k , which is defined by

$$\mathcal{T}_k(M) = \int_{U \in \mathcal{U}_N} U^{\otimes k} M (U^\dagger)^{\otimes k}. \quad (34)$$

(Throughout the paper the integration measure dV will be the Haar measure.) This map is the “twirling” operation[2]. Since \mathcal{T}_k is a Hermitian map and $\mathcal{T}_k(\mathcal{T}_k(M)) = \mathcal{T}_k(M)$, the map $\mathcal{T}_k(M)$ has all eigenvalues equal to zero or unity.

For $\pi \in \mathcal{S}_k$, we define the $N^k \times N^k$ matrix $\mathbf{P}_N(\pi)$ is defined to be

$$\mathbf{P}_N(\pi) = \sum_{i_1=1}^N \cdots \sum_{i_k=1}^N |i_1, \dots, i_N\rangle \langle i_{\pi(1)}, \dots, i_{\pi(N)}|.$$

Since $\mathbf{P}_N(\pi)$ commutes with any matrix of the form $U^{\otimes k}$, it follows that $\mathcal{T}_k(\mathbf{P}_N(\pi)) = \mathcal{E}_k(\mathbf{P}_N(\pi)) = \mathbf{P}_N(\pi)$ for any π . We claim that the $\mathbf{P}_N(\pi)$ (and their linear combinations) constitute all of the unit eigenvalues of \mathcal{E}_k . This fact follows from Schur-Weyl duality, and specifically Thm 3.3.8 of [24] which states that $\mathcal{T}_k(M) = M$ if and only if M is a linear combination of $\mathbf{P}_N(\pi)$ operators. Thus $F_k^N = \dim \text{Span}\{\mathbf{P}_N(\pi) : \pi \in \mathcal{S}_k\}$.

An important special case is when $N \geq k$. In this case, the set $\{\mathbf{P}_N(\pi) | 1, 2, \dots, k\} : \pi \in \mathcal{S}_k\}$ is linearly independent, which implies that $\{\mathbf{P}_N(\pi) : \pi \in \mathcal{S}_k\}$ is linearly independent and thus that $F_k^N = k!$.

In the quantum case, tensor product expanders give us a way to approximate the twirling operator \mathcal{T}_k of [2]. This is because

$$\|\mathcal{E}_k^m - \mathcal{T}_k\|_\infty \leq \lambda^m, \quad (35)$$

so whenever $\lambda < 1$, $\mathcal{E}_k^\infty = \mathcal{T}_k$. Let us consider various other possibilities for implementing twirling as a sum of different unitary transformations: one approach to exactly implementing the twirling operation is to use t -designs[1], but the

² One can slightly generalize this by defining a set of unitaries and a set of associated probabilities to be a tensor product expander; however in this paper we consider applying each unitary with equal probability summing to unity.

number of unitaries that must be implemented in this case grows with N . Another approach was discussed in [20], which avoids having the number of unitaries grow in N , but requires the ability to implement a number of unitaries growing linearly in the logarithm of the error of the approximation. In contrast, tensor product expanders require only the ability to implement a constant number of unitaries to get arbitrarily good approximations. This is a definite advantage; however, in practice, our construction of tensor product expanders here, which relies on the ability to construct random unitary operations, probably cannot be efficiently implemented using gates; instead, we would like to efficiently implement a deterministically constructed tensor product expander. This raises the interesting question of whether the constructions of [5] can lead to tensor product expanders also.

The limit of large k : The situation when k is large has some similarities to the classical case. It still holds that any (N, D, λ, k) quantum tensor product expander is also a (N, D, λ, k') quantum tensor product expander for all $k' \leq k$. In particular, if a set of unitaries forms a (N, D, λ, ∞) quantum tensor product expander then it is also a (N, D, λ, k) quantum tensor product expander for any finite k . This is equivalent to generating a Cayley graph expander on \mathcal{U}_N . One difference between the quantum and classical cases is that there is no upper bound to the size of irreps of \mathcal{U}_N , like there is for \mathcal{S}_N .

Note that constant degree Cayley graph expanders are known for \mathcal{U}_2 ; indeed, choosing the matrices at random will yield an expander with probability one [26]. However, no proof of this fact is known for $N > 2$.

B. Solovay-Kitaev gate approximation

One application of tensor product expanders is to the problem of approximating an arbitrary $V \in \mathcal{U}_N$ with a string of gates from a fixed universal set $\{U(1), \dots, U(D)\}$. The fact that $\{U(1), \dots, U(D)\}$ is universal means that $\langle U(1), \dots, U(D) \rangle$ is dense in \mathcal{U}_N (optionally neglecting an overall phase). This means that for any $V \in \mathcal{U}_N$ and any $\epsilon > 0$, there exists a string s_1, \dots, s_m such that $U(s_1)U(s_2) \cdots U(s_m)$ is within a distance ϵ of V . Often we also want to know (a) how quickly m grows with $1/\epsilon$ and (b) how long it takes to find s_1, \dots, s_m . When $\{U(1), \dots, U(D)\}$ contain their own inverses, the Solovay-Kitaev theorem [21] gives a $\text{poly} \log(1/\epsilon)$ time (for fixed N) algorithm to find an ϵ -approximation with $m = O(\log^{3+o(1)}(1/\epsilon))$. Very little is known in the case without access to inverses, except that $U(s)^\dagger$ can be simulated to error ϵ using $O(1/\epsilon^{N^2})$ applications of $U(s)$, meaning that the Solovay-Kitaev construction can be used with this amount of overhead.

Turning to lower bounds, observe a ball of radius ϵ in \mathcal{U}_N has volume $\Theta(\epsilon^{N^2})$. This implies that to approximate all strings to within error ϵ requires $\Omega((1/\epsilon)^{N^2})$ different unitaries, or equivalently a $\Omega(N^2 \log 1/\epsilon)$ string length. A long-standing open question is whether the Solovay-Kitaev approximation can in general be improved to use the optimal $O(\log 1/\epsilon)$ number of gates. Such optimally short approximations are known to exist whenever a particular random walk on \mathcal{U}_N has a gap [22]: specifically, the walk consisting of multiplying by $U(s)$ for s randomly chosen from $1, \dots, D$. For \mathcal{U}_2 , it was recently proven that generic $U(1), \dots, U(D)$ are gapped [23] and thus yield short approximating strings. However, the situation for \mathcal{U}_N for $N > 2$ remains open.

In this section we will prove that when k is sufficiently large, unitaries forming k -tensor product expanders yield optimal $O(N^2 \log 1/\epsilon)$ -length ϵ -approximations for any gate in \mathcal{U}_N .

Theorem 5 . *Suppose $\{U(1), \dots, U(D)\}$ form a k -tensor product expander with gap $1 - \lambda$ for $k \gg \frac{N^3 \log^2(1/\epsilon)}{\epsilon}$. Then for any $V \in \mathcal{U}_N$ there exists a string $s_1, \dots, s_m \in \{1, \dots, D\}$ with $m = O(N^2 \log_{1/\lambda}(1/\epsilon))$ and $d(V, U(s_1)U(s_2) \cdots U(s_m)) \leq \epsilon$.*

Here we define the distance between two unitaries $d(U, V)$ by

$$d(U, V) = \min_{\phi \in [0, 2\pi]} \|U - e^{i\phi} V\|_2 = 2N - 2|\text{tr } U^\dagger V|,$$

so that it ignores overall phase.

The main result from [22] can be thought of a weaker version of Theorem 5: it requires $k = \infty$ to achieve the same conclusion. Unfortunately, Theorem 6 only shows that generic sets of unitaries are k -tensor product expanders for $k \sim N^{1/6}/\log(N)$. Thus, at present the existence of expanders satisfying the assumptions of Theorem 5 is a nontrivial conjecture. It is possible that there exists some strengthening of the results of Theorem 6 which will allow us to show that generic unitaries fulfill the assumptions of Theorem 5.

Proof of Theorem 5: Let $|\Phi\rangle = \frac{1}{\sqrt{N}} \sum_{i=1}^N |i\rangle|i\rangle$ be the maximally entangled state on $\mathbb{C}^N \otimes \mathbb{C}^N$. Define $\rho(U) = [(U \otimes I)\Phi(U^\dagger \otimes I)]^{\otimes k}$. Observe that

$$\text{tr } \rho(U)\rho(V) = |\text{tr } U^\dagger V|^{2k} / N^{2k} = \left(1 - \frac{d(U, V)}{2N}\right)^{2k} \quad (36)$$

Let $B_{\epsilon/3}$ be the ball of radius $\epsilon/3$ around the identity: $B_{\epsilon/3} = \{U | d(U, I) \leq \epsilon/3\}$. Let $\text{Vol}(\epsilon/3)$ denote the volume of $B_{\epsilon/3} = \mathcal{O}((\epsilon/3)^{N^2})$. Define

$$\rho_\epsilon(U) = \frac{1}{\text{Vol}(\epsilon/3)} \int_{V \in B_{\epsilon/3}} \rho(VU) dV, \quad (37)$$

Similarly we define

$$\rho_H = \int_{V \in \mathcal{U}_N} \rho(V) dV. \quad (38)$$

These states are normalized so that $\text{tr } \rho_\epsilon(U) = \text{tr } \rho_H = 1$. Since $\rho(V) \geq 0$ for all V , we have the operator inequality $\rho_\epsilon(U) \leq \rho_H / \text{Vol}(\epsilon/3)$ for any U . Also observe that $\rho_H = (\mathcal{T}_k \otimes \text{id}_N^{\otimes k})(\rho(U))$ for any U , where id_N denotes the identity operation on $N \times N$ density matrices.

We will find it convenient to think of density matrices as vectors with the Hilbert-Schmidt inner product $\langle A, B \rangle = \text{tr } A^\dagger B$. In this picture \mathcal{T}_k is a projector, and so

$$\text{tr } \rho_\epsilon(U) \rho_H = \text{tr } \rho_\epsilon(U) (\mathcal{T}_k \otimes \text{id}_N^{\otimes k})(\rho_\epsilon(U)) = \text{tr } \rho_H^2.$$

To bound $\text{tr } \rho_H^2$, observe that the support of ρ_H lies within $\text{Span}\{|\psi\rangle^{\otimes k} : |\psi\rangle \in \mathbb{C}^{N^2}\}$, which (according to [24, 25]) has dimension $\binom{N^2+k-1}{N^2} = k(k+1) \cdots (k+N^2-1)/N^2! \leq k^{N^2}$. Thus $\text{tr } \rho_H^2 \geq k^{-N^2}$.

Now we use the fact that $\|\mathcal{E}_k^m - \mathcal{T}_k\|_\infty \leq \lambda^m$ together with Cauchy-Schwartz to bound

$$\text{tr } \rho_\epsilon(I) \mathcal{E}^m(\rho_\epsilon(U)) \geq \text{tr } \rho_\epsilon(I) \rho_H - \lambda^m \text{tr } \rho_\epsilon(I)^2 \geq \text{tr } \rho_H^2 \left(1 - \frac{\lambda^m}{\text{Vol}(\epsilon/3)^2}\right) \geq \frac{1}{2} \text{tr } \rho_H^2 \geq \frac{1}{2k^{N^2}}, \quad (39)$$

where in the second-to-last step we have assumed $m \geq \log(2/\text{Vol}(\epsilon/3)^2)/\log(1/\lambda) = \mathcal{O}(N^2 \log_{1/\lambda}(1/\epsilon))$.

On the other hand, if there is no string s_1, \dots, s_m such that $d(U(s_1)U(s_2)\dots U(s_m), U) \leq \epsilon$, then

$$\text{tr } \rho_\epsilon(I) \mathcal{E}^m(\rho_\epsilon(U)) \leq \left(1 - \frac{\epsilon}{6N}\right)^{2k} \leq e^{-\frac{k\epsilon}{3N}}. \quad (40)$$

If $k/\log k \gg N^3/\epsilon$ then (39) and (40) cannot simultaneously hold. Therefore there must exist at least one string s_1, \dots, s_m for which $d(U(s_1)U(s_2)\dots U(s_m), U) \leq \epsilon$. ■

C. Trace Method and Schwinger-Dyson Equations

The next three sections are devoted to the expansion properties of randomly chosen unitaries. Recall that we would like to construct a quantum tensor product expander by randomly choosing $U(1), \dots, U(D) \in \mathcal{U}_N$. There are two cases. In the non-Hermitian case, the unitary matrices $U(s)$ are chosen independently with the Haar measure. In the Hermitian case, D is even and the unitary matrices $U(s)$ for $s = 1, \dots, D/2$ are chosen independently with the Haar measure and $U(s + D/2) = U(s)^\dagger$, so that \mathcal{E}_k is a Hermitian operator. We focus on the Hermitian case, and the techniques can be readily extended to cover the non-Hermitian case. Our main result is that for random $U(s)$, with high probability we do indeed get a tensor product expander:

Theorem 6 . *Let $\{U(1), \dots, U(D/2)\}$ be chosen randomly with the Haar measure from the unitary group \mathcal{U}_N , and let $U(s + D/2) = U(s)^\dagger$. Let $k \leq \mathcal{O}(N^{1/6}/\log(N))$ and let λ denote the $F_k^N + 1^{\text{st}}$ eigenvalue of \mathcal{E}_k as defined in (32). Then, for any $c > 1$,*

$$\Pr \left[\lambda \geq c(1 + \mathcal{O}(k \log(N) N^{-1/6})) \lambda_H \right] \leq c^{-(1/4k)N^{1/6}}, \quad (41)$$

where λ_H depends on D and is given in Eq. (14).

We use a trace method to bound the eigenvalues of $\mathcal{E}_k(M)$. We have

$$\sum_{i_1, i_2, \dots, i_k} \sum_{j_1, j_2, \dots, j_k} \left(M(i_1, j_1) \otimes M(i_2, j_2) \otimes \dots \otimes M(i_k, j_k), \mathcal{E}_k^m(M(i_1, j_1) \otimes M(i_2, j_2) \otimes \dots \otimes M(i_k, j_k)) \right) = \sum_{a=1}^{N^{2k}} |\lambda_a|^m \geq k! + |\lambda|^m, \quad (42)$$

where we pick m to be an even integer. We will derive bounds on the expectation value of the trace to bound the expectation of $|\lambda|^m$. Eq. (42) can be re-written as

$$k! + |\lambda|^m \leq \left(\frac{1}{D}\right)^m \sum_{s_1=1}^D \sum_{s_2=1}^D \dots \sum_{s_m=1}^D \text{tr}(U(s_m + D/2) \dots U(s_2 + D/2) U(s_1 + D/2))^k \text{tr}(U(s_1) U(s_2) \dots U(s_m))^k. \quad (43)$$

Let $\mathbb{E}[\dots]$ denote the average over the unitary group. Averaging Eq. (43) we find

$$E_{1,k} \equiv \left(\frac{1}{D}\right)^m \sum_{s_1=1}^D \sum_{s_2=1}^D \dots \sum_{s_m=1}^D E_{0,k}(s_1, \dots, s_m) \geq k! + \mathbb{E}[|\lambda|^m], \quad (44)$$

$$\begin{aligned} E_{0,k}(s_1, \dots, s_m) &\equiv \mathbb{E}[\text{tr}(U^\dagger(s_m) \dots U^\dagger(s_2) U^\dagger(s_1))^k \text{tr}(U(s_1) U(s_2) \dots U(s_m))^k] \\ &= \mathbb{E}[\text{tr}(U(s_m + D/2) \dots U(s_2 + D/2) U(s_1 + D/2))^k \text{tr}(U(s_1) U(s_2) \dots U(s_m))^k]. \end{aligned} \quad (45)$$

As in [19], we write the average in Eq. (45) as an average of the form

$$\mathbb{E}[L_1 L_2 \dots L_c], \quad (46)$$

where

$$L_1 = \text{tr}(U(s_{1,1}) U(s_{1,2}) \dots U(s_{1,m_1})), \quad L_2 = \text{tr}(U(s_{2,1}) U(s_{2,2}) \dots U(s_{2,m_2})), \quad \dots \quad (47)$$

Here we have an average of c traces, each of which is a product of some number of unitary matrices. In particular, Eq. (45) has $c = 2k$, with $L_1 = L_2 = \dots = L_k = L_{k+1}^\dagger = \dots = L_{2k}^\dagger$.

The Schwinger-Dyson equations for a product of this form are [19]:

$$\begin{aligned} &\mathbb{E}[\text{tr}(U(s_{1,1}) U(s_{1,2}) \dots U(s_{1,m_1})) L_2 \dots L_c] \\ &= -\frac{1}{N} \sum_{j=2}^{m_1} \delta_{s_{1,1}, s_{1,j}} \mathbb{E}[\text{tr}(U(s_{1,1}) \dots U(s_{1,j-1})) \text{tr}(U(s_{1,j}) \dots U(s_{1,m_1})) L_2 \dots L_c] \\ &+ \frac{1}{N} \sum_{j=2}^{m_1} \delta_{s_{1,1}, s_{1,j+D/2}} \mathbb{E}[\text{tr}(U(s_{1,2}) \dots U(s_{1,j-1})) \text{tr}(U(s_{j+1,1}) \dots U(s_{1,m_1})) L_2 \dots L_c] \\ &- \frac{1}{N} \sum_{l=2}^c \sum_{j=1}^{m_l} \delta_{s_{1,1}, s_{l,j}} \mathbb{E}[\text{tr}(U(s_{1,1}) \dots U(s_{1,m_1}) U(s_{l,j}) U(s_{l,j+1}) \dots U(s_{l,j-1})) L_2 \dots L_{l-1} L_{l+1} \dots L_c] \\ &+ \frac{1}{N} \sum_{l=2}^c \sum_{j=1}^{m_l} \delta_{s_{1,1}, s_{l,j+D/2}} \mathbb{E}[\text{tr}(U(s_{1,2}) \dots U(s_{1,m_1}) U(s_{l,j+1}) U(s_{l,j+2}) \dots U(s_{l,j-1})) L_2 \dots L_{l-1} L_{l+1} \dots L_c]. \end{aligned} \quad (48)$$

Note that in the above equation an expression like $U(s_{l,j+1}) U(s_{l,j+2}) \dots U(s_{l,j-1})$ means $U(s_{l,j+1}) U(s_{l,j+2}) \dots U(s_{l,m_l}) U(s_{l,1}) U(s_{l,2}) \dots U(s_{l,j-1})$.

Our general algorithm for reducing traces starts by canceling all pairs of matrices $U(s)U(s + D/2)$ appearing successively in the same trace, and replacing $\text{tr}(\mathbb{1})$ by N . We then apply Eq. (48), repeating the cancellation of successive $U(s)U(s + D/2)$ and replacement of $\text{tr}(\mathbb{1})$ by N on each iteration. A term terminates at a given level n if there are no matrices left after n iterations.

Let m_1^0 be the length of the trace after canceling successive $U(s)U(s + D/2)$ before any iterations; on every successive iteration, the length of the first trace, m_1 , is bounded by m_1^0 . As in [19], the number of different choices of s_1, \dots, s_m which give rise to a given m_1^0 is bounded by

$$(D-1)^{m_1^0/2} (D-1)^{m/2} 2^m. \quad (49)$$

This number is equal to D^m times the probability that a random walker on a Cayley tree arrives at a distance m_1^0 from the starting point after a walk of m steps. This number is independent of the particular values of $s_{1,1}, \dots, s_{1,m_1^0}$. There are $[D/(D-1)](D-1)^{m_1^0}$ different possible values of $s_{1,1}, \dots, s_{1,m_1^0}$ and therefore the total number of choices of s_1, \dots, s_m which give rise to a given choice of $s_{1,1}, \dots, s_{1,m_1^0}$ is bounded by

$$\frac{D-1}{D} \left(\frac{1}{\sqrt{D-1}} \right)^{m_1^0} (D-1)^{m/2} 2^m. \quad (50)$$

The number of terms terminating at the n^{th} level is bounded by

$$(2km - 1)^n. \quad (51)$$

To see this, note that at each iteration of the Schwinger-Dyson equation, the number of terms on the right-hand side is bounded by the number of matrices on the left-hand side minus one. Initially, there are $2km$ matrices, and this number does not increase under Eq. (48).

We can estimate the value of a term which terminates at a given level $n > 1$ as follows. First, there is a sign equal to plus or minus 1. Next, there is a factor of $(1/N)^n$. Finally, there is a factor of N for each trace of the form $\text{tr}(\mathbb{1})$ that appeared in this process. Suppose there are p such traces, giving a factor of N^p . How big can p be? Initially we have $c = 2k$ different traces. The given term at level n arose from a specific choice of terms on the right-hand side of Eq. (48) on the first iteration. This specific choice has k_1 different traces in it, with k_1 equal to either $k - 1$ or $k + 1$. After the second iteration there are k_2 traces, then k_3 , and so on. The number of traces k_2, k_3, \dots can be determined as follows: an application of Eq. (48) may increase the number of traces by one if the term arises from the first or second line on the right-hand side, or may decrease the number of traces by one if the term arises from the third or fourth line on the right-hand side of Eq. (48). Next, some of the traces may be trivial, being equal to $\text{tr}(\mathbb{1})$. In the event that the term arose from the first, second, or third line of Eq. (48) it is not possible for any of the traces to be trivial, under the assumption that any repetitions of the form $U(s)U(s + D/2)$ have been previously replaced by $\mathbb{1}$ in the trace on the left-hand side of the equation. However, in the event that the term arose from the fourth line, then it is possible for one of the traces to be trivial, increasing p by one. Thus, for each $b \leq n$, $k_b - k_{b-1}$ is equal to either $+1, -1$, or -2 . Let q be equal to the number of times the first or second line was used from Eq. (48) and $n - q$ equal the number of times the third or fourth line was used. Then, in order for all traces to be trivial in this particular term resulting from n iterations of Eq. (48),

$$2k + q - (n - q) - p = 0. \quad (52)$$

Also, since p can only increase when a term from the fourth line is used,

$$p \leq n - q. \quad (53)$$

Thus,

$$p \leq \lfloor (2k + n)/3 \rfloor. \quad (54)$$

Therefore, the value of a term terminating at the n^{th} level, $n > 0$, is bounded in absolute value by

$$N^{\lfloor (2k+n)/3 \rfloor - n}. \quad (55)$$

Note that if $m_1^0 > 0$ then there are no terms terminating at level n with $n < k$, so for $m_1^0 = 0$, the trace is equal to N^{2k} , while for $m_1^0 > 0$, the terms are bound in absolute value by N^0 (this bound is only reached if $k = n$).

Eq. (48) generates an infinite series, whose n^{th} term is the sum of all terms terminating at level n . As in [19], this series is absolutely convergent for $2km < N$. In fact, the following stronger claim holds: Eq. (48) generates an absolutely convergent series for $2km - 1 < N$ which converges to the expectation value of the trace. To see this, note that the value p above, the number of traces of $\mathbb{1}$, is always bounded by $2km$. Thus, the value of a term terminating at the n^{th} level is bounded by

$$N^{2km} N^{-n}. \quad (56)$$

Depending on n , sometimes (55) gives a better bound and sometimes (56) gives a better bound, but to estimate convergence we will use (56). Eq. (51) shows that the number of terms terminating at level n is bounded by $(2km - 1)^n$. Thus, the absolute value of the sum of terms terminating at level n is bounded by $N^{2km} ((2km - 1)/N)^n$, and so for $2km - 1 < N$, the series is absolutely convergent. Further, a term which has not terminated at the n^{th} level contains at most $2km$ traces in it, and hence is bounded in absolute value by $N^{2km} (1/N)^n$. Therefore, the sum of all terms which have not terminated at the n^{th} level is also bounded by $N^{2km} ((2km - 1)/N)^n$, and hence for $2km - 1 < N$ the series converges to the average of the trace.

D. Example

We now work out a simple example to give some idea of the use of the Schwinger-Dyson equations. This example will also be used later in the idea of “complete rung cancellation” and gives intuition behind the claim that for $N \geq k$

we have $k!$ eigenvalues equal to unity. Let the matrix X be chosen from the unitary group with the Haar measure and evaluate the expectation value for $N \geq k$

$$\mathbb{E}[\left(\text{tr}(X)\text{tr}(X^\dagger)\right)^k]. \quad (57)$$

For $k = 1$, a single application of Eq. (48) shows that this is equal to unity. For $k = 2$, we find

$$\begin{aligned} \mathbb{E}\left[\left(\text{tr}(X)\text{tr}(X^\dagger)\right)^2\right] &= 2\mathbb{E}\left[\left(\text{tr}(X)\text{tr}(X^\dagger)\right)\right] - (1/N)\mathbb{E}\left[\text{tr}(XX)\left(\text{tr}(X^\dagger)^2\right)\right] \\ &= 2 + (1/N)^2\mathbb{E}\left[\left(\text{tr}(X)\text{tr}(X^\dagger)\right)^2\right] - 2(1/N)^2\mathbb{E}\left[\text{tr}(X)\text{tr}(X^\dagger)\right] \\ &= 2 + (1/N)^2\mathbb{E}\left[\left(\text{tr}(X)\text{tr}(X^\dagger)\right)^2\right] - 2(1/N)^2. \end{aligned} \quad (58)$$

For $N \geq 2$, this shows that $\mathbb{E}\left[\left(\text{tr}(X)\text{tr}(X^\dagger)\right)^2\right] = 2$.

It is interesting to see what happens to the expectation value in Eq. (58) for $N = 1, k = 2$. Then, the last line Eq. (58) gives simply $\mathbb{E}\left[\left(\text{tr}(X)\text{tr}(X^\dagger)\right)^2\right] = \mathbb{E}\left[\left(\text{tr}(X)\text{tr}(X^\dagger)\right)^2\right]$, giving no information about the trace. For general N , the sum of terms terminating at level 1 is equal to zero, while the sum of terms terminating at levels 2, 3, 4, 5, 6... is equal to 2, $-2/N, 2/N, -2/N^2, 2/N^2, \dots$ respectively. Thus, we do not have a convergent series for $N = 1, k = 2$.

Up to now we have considered the series whose n^{th} term is the sum of terms terminating at a given level n . We now consider instead the expectation value of Eq. (57) as a series in $1/N$. For $N \geq k$, this series is again absolutely convergent to the desired expectation value. It is easy to see that for arbitrary k , and for $N \gg k$, the expectation value (57) is equal to $k! + \mathcal{O}(1/N)$, as there are $k!$ terms which terminate at level k . We now show that for $N \geq k$, the expectation value (57) is equal to $k!$ exactly. Note that the expectation value in Eq. (57) is equal to the trace of the map \mathcal{T}_k (defined in (34))

Thus, the trace of the map $\mathcal{T}_k(M)$ is equal to the number of unit eigenvalues of $\mathcal{T}_k(M)$. For $N \geq k$ the trace of this map can then be written as the sum of an infinite series in $1/N$, and using the fact that the number of unit eigenvalues is equal to an integer for all integer N , we find that all terms in the series in $1/N$, beyond the term of order N^0 , must vanish exactly (the calculation above represents an explicit check of this for $k = 2$ and it may be readily verified for any k). Thus, for all $N \geq k$, the expectation value of Eq. (57) is equal to $k!$. This gives an alternate proof that $F_k^N = k!$ when $N \geq k$.

E. Counting and Main Result

In this section we prove a bound on the expectation value of the sum in Eq. (44), which will give us a bound on the expectation value of the m^{th} power of λ , proving the theorem. The next three paragraphs are devoted to outlining the basic idea of the proof, before beginning the technical details.

The basic idea of the proof is to prove the bound on the sum by proving a bound on the number of different choices of s_1, \dots, s_m such that, when the resulting trace is evaluated using the Schwinger-Dyson equations, there is a term which terminates at level n , for any given n . We give this bound on the number of choices of s_1, \dots, s_m in Eq. (61). We then combine this bound with a bound on the contribution to the trace of terms which terminate at level n . The idea is that there are a only small number of choices of s_1, \dots, s_m which produce terms which terminates at a small level n , and while there are a large number of choices of s_1, \dots, s_m which produce terms which terminate at high levels, such terms are small.

One technical caveat in this work is that for *any* choice of s_1, \dots, s_m there will be certain terms which terminate at a low level n . These are terms in which we use the Schwinger-Dyson equations to contract $U(s_i)$ in one trace with $U(s_i)^\dagger$ in a different trace. If for some i , we contract all unitaries $U(s_i)$ in this way, we have what is called a “complete rung cancellation” below. We consider such terms separately, and they are responsible for producing the leading order expectation value of the trace in $1/N$: these terms sum to give a contribution $k!$ to the expectation value of the trace, precisely corresponding to the expectation we expect from the unit eigenvalues.

Ignoring those terms with complete rung cancellations, we see that a term in the Schwinger-Dyson equations must involve contracting $U(s_i)$ with $U(s_j)$ or $U(s_j)^\dagger$ for some $i \neq j$. Such terms involve constraints: such a term would require that either $s_i = s_j + D/2$ or $s_i = s_j$. In order for such a term to terminate at a low level, there must be many such constraints, and this is why there are only a few choices of s_1, \dots, s_m which produce terms which terminate at low levels. To show precisely that there are only a few such choices of s_1, \dots, s_m , we follow a different strategy. To

explain this strategy, suppose you knew a choice of s_1, \dots, s_m which gave rise to a term which terminated at some level n and you were given the task of explaining to someone which choice of s_1, \dots, s_m you used. One way to do this would be to simply list the m different values of s . This would require communicating $\log_2(D^m)$ bits. We instead show how to uniquely specify the choices of s_1, \dots, s_m in a different way, by specifying most of the choices of s_1, \dots, s_m by describing which cancellations were used. For small n , this will allow one to communicate the specific choice of s_1, \dots, s_m in much shorter way, thus implying that there are only a few choices of s_1, \dots, s_m which produce the desired term terminating at level n . We now put this idea into practice.

On a given iteration of the Schwinger-Dyson equations, we go from a product of c traces to a product of $c + 1$, $c - 1$, or $c - 2$ traces. As in [19], we keep track of how the matrices move under this iteration process using a function $f_n((l, i))$ from pairs of integers to pairs of integers. We say that the matrix $U(s_{l,i})$ in the given product of traces, $L_1 L_2 \dots L_c$, is in position (l, i) . Let us consider the case of a term on the first line, where c increases by one. Then, for any given j in the sum on the first line, we say that the matrix in position $(1, i)$, for $i < j$ on the $n + 1^{\text{st}}$ iteration corresponds to the matrix in position $(1, i)$ on the n^{th} iteration, and so $f_n((1, i)) = (1, i)$, while the matrix in position $(2, i)$ on the $n + 1^{\text{st}}$ iteration corresponds to the matrix in position $(1, i + j - 1)$ on the n^{th} iteration, so $f_n((1, i + j - 1)) = (2, i)$. The matrix in position (l, i) , for $2 < l \leq k + 1$ on the $n + 1^{\text{st}}$ iteration corresponds to the matrix $(l - 1, i)$ on the n^{th} iteration, so $f_n(l - 1, i) = (l, i)$. We follow a similar procedure for the other lines of Eq. (48) and if there are cancellations, we keep track of how the matrix moves under the cancellations.

We then keep track of which matrix after n iterations corresponds to a given matrix before any iterations, by defining $F_n((l, i)) = f_n(f_{n-1}(\dots f_1((l, i)))$ for $l = 1, 2, \dots, 2k$. Let us say that the matrix at position (l, i) is “trivially moved” under the n^{th} iteration of the Schwinger-Dyson equations if it is not in either position $(1, 1)$ or position $(1, j)$ using a term on the first or second line, or in either position $(1, 1)$ or position (l, j) using a term from the third or fourth line. If a matrix is not trivially moved, and the matrix is not in position $(1, 1)$, then the Schwinger-Dyson equations imply a relation between $s_{l,i}$ and $s_{1,1}$.

A given term in Eq. (48) arises from a given choice of (l, j) : for a term on the first or second line let us say $l = 1$. Let $(1, 1) = F_n(l_0, j_0)$ and let $(l, j) = F_n(l'_0, j'_0)$. If a matrix is *not* trivially moved under on the n^{th} iteration then there are two cases: (1) either $l_0 \leq k$ and $l'_0 \leq k$ or $l_0 > k$ and $l'_0 > k$. That is, either both matrices appeared in one of the first k traces, which are traces of products of conjugates of unitaries, or both matrices appeared in one of the last k traces, which are traces of unitaries. Or, case (2): $l_0 \leq k$ and $l'_0 > k$ or $l_0 > k$ and $l'_0 \leq k$. That is, one matrix was in one of the first k traces and the other was in one of the last k traces. We then break the first case into two sub-cases: (a), $j_0 = j'_0$ or (b), $j_0 \neq j'_0$. We also break the second case into two sub-cases: (a), $j_0 = m_1 + 1 - j'_0$ or (b), $j_0 \neq m_1 + 1 - j'_0$. In case 1a both matrices are unitary matrices $U(s_{1,j_0})$ or both are $U(s_{1,j_0})^\dagger$ and in case 2a, one matrix is $U(s_{1,j_0})$ and the other is $U(s_{1,j_0})^\dagger$. In case 1b, we know that $s_{1,j_0} = s_{1,j'_0}$ for $j_0 \neq j'_0$ while in case 2b we know that $s_{1,j_0} = s_{1,j'_0} + D/2$ for $j_0 \neq j'_0$. Thus, in case 1b or 2b the term in the Schwinger-Dyson equation implies some constraint about the choice of $s_{1,j}$. To illustrate these different cases, consider the example (58): the first term on the right-hand side of the top line is an example of case 2a, while the second term on the same line is an example of case 1a.

Consider a given j ; if on some iteration and for some l the matrix which was originally in position (l, j) is not trivially moved and we have case 1b or 2b, then we can identify some k such that either $s_{1,j} = s_{1,k}$ or $s_{1,j} = s_{1,k} + D/2$. Let us write $k = \tau(j)$ in both cases, for some function $\tau(j)$. We define a term to have a “complete rung cancellation of matrix j ” if it is not possible to identify such a k for the given j . We claim that the sum of all terms with a complete rung cancellation of matrix i is equal to $k!$ so long as $k \leq N$. To show this, consider the product of traces

$$\text{tr}(U(s_m + D/2) \dots U(s_{i+1} + D/2) X^\dagger U(s_{i-1} + D/2) \dots U(s_1 + D/2))^k \text{tr}(U(s_1) \dots U(s_{i-i}) X U(s_{i+1}) \dots U(s_m)^k, \quad (59)$$

where X is some arbitrary unitary matrix. Averaging this trace over all unitary matrices $U(s)$ and over all unitary matrices X with the Haar measure, we find that the trace is equal to $k!$: this can be established by applying Eq. (48) to this trace, and always cyclically permuting the trace so that X is in the first position. This calculation is very similar to the example calculation (57) above. However, applying the Schwinger-Dyson equations to the trace (59) without first applying the cyclic permutation generates precisely the sum of terms mentioned above, those in which there is a complete rung cancellation of matrix i . Thus, this sum of terms equals $k!$. We further claim that for any given i_1, i_2, \dots, i_d , the sum of all terms with complete rung cancellations of matrices i_1, i_2, \dots, i_d is equal to $k!$, as may be shown by considering a trace in which matrices $U(s_{i_1}), U(s_{i_2}), \dots$ are replaced by X_1, X_2, \dots , and the trace is averaged over the different X_1, X_2, \dots . Then, using the inclusion-exclusion principle, the sum of terms in which for no i is there a complete rung cancellation of matrix i is equal to the sum of all terms minus $k!$. So, we now focus on the sum of terms with no complete rung cancellations, which we define to be $E'_{0,k}(s_1, \dots, s_m)$; if a given choice of s_1, \dots, s_m gives rise to a term which terminates at level n with no complete rung cancellations, then it is possible to identify a $\tau(i)$ for each i .

We now follow the same approach as in [19] to bound the number of choices of $s_1, \dots, s_{m_1^0}$ which can produce a term which terminates at a level n with no complete rung cancellations. Given the sequence of choices of terms on the

right-hand side of the Schwinger-Dyson equation (48), as well as knowledge of which cancellations occurred at each iteration, we know the function $\tau(i)$, and given this function $\tau(i)$ there are now only at most $[D/(D-1)](D-1)^{m_1^0/2}$ possible values of $s_{1,1}, \dots, s_{1,m_1^0}$. Thus, the total number of choices of $s_1, \dots, s_{m_1^0}$ which can produce a term which terminates at level n is bounded by the number of possible choices of terms and cancellations in the Schwinger-Dyson equation (48) at each of the n iterations multiplied by $[D/(D-1)](D-1)^{m_1^0/2}$. At each iteration of the Schwinger-Dyson equations, we make a particular choice of l, j at each level, which requires specifying one particular matrix out of all the matrices on the right-hand side; there are at most $2m_1k - 1$ matrices on the right-hand side, so there are at most $2m_1k - 1$ choices (in [19], the slightly worse bound $(2m_1k - 1)^2$ was found; we tighten the bound here). At each such iteration of the Schwinger-Dyson equations, there may be cancellations in two different traces if the term came from the second line of Eq. (48), with at most m_1 cancellations in each trace, or cancellations in two different places of a single trace, if the term came from the fourth line of Eq. (48), with at most m_1 cancellations in each place. Let us call the number of cancellations c_1, c_2 with $0 \leq c_1 \leq m_1$ and $0 \leq c_2 \leq m_1$. Then, by specifying l, j, c_1, c_2 for each iteration, we succeed in fully specifying how the matrices move under the n iterations of the Schwinger-Dyson equation; this requires specifying n numbers ranging from $1 \dots 2km_1 - 1$, and $2n$ numbers ranging from $0 \dots m_1$.

Thus, there are at most

$$[D/(D-1)](D-1)^{m_1^0/2}(2km_1^0 - 1)^n(m_1^0 + 1)^{2n} \leq [D/(D-1)](D-1)^{m_1^0/2}(2km_1^0)^{3n} \quad (60)$$

choices of $s_1, \dots, s_{m_1^0}$ which can produce a term which terminates at level n . Using Eq. (50), the number of choices of s_1, \dots, s_m which can produce a term which terminates at level n is at most

$$\sum_{m_1^0=0}^m (D-1)^{m/2} 2^m (2km_1^0)^{3n} \leq (D-1)^{m/2} 2^m \frac{(2km+1)^{3n+1}}{3n+1}. \quad (61)$$

For any s_1, \dots, s_m , we define $n_{\min}(s_1, \dots, s_m)$ to be the smallest level at which a term terminates with no complete rung cancellations. The sum of terms with $m_1^0 = 0$, which is the same as the sum of terms with $n_{\min} = 0$, is bounded by

$$N^{2k} D^{-m} (D-1)^{m/2} 2^m = N^2 \lambda_H^m. \quad (62)$$

Thus, we re-write the sum in Eq. (44) as

$$E_{1,k} \leq k! + N^{2k} N^2 \lambda_H^m \left(\frac{1}{D}\right)^m \sum_{n=k}^{\infty} \sum_{s_1=1}^D \sum_{s_2=1}^D \dots \sum_{s_m=1}^D \delta_{n_{\min}(s_1, \dots, s_m), n} E'_{0,k}(s_1, \dots, s_m). \quad (63)$$

Therefore, for any s_1, \dots, s_m with $n_{\min} > 0$,

$$\begin{aligned} E'_{0,k}(s_1, \dots, s_m) &\leq \sum_{n \geq n_{\min}(s_1, \dots, s_m)} N^{2(k-n)/3} (2km-1)^n \\ &= N^{2k/3} \frac{[N^{-2/3}(2km-1)]^{n_{\min}}}{1 - N^{-2/3}(2km-1)}. \end{aligned} \quad (64)$$

From Eqs. (61,63,64),

$$\begin{aligned} E_{1,k} &\leq k! + \lambda_H^m \left\{ N^2 + N^{2k/3} \sum_{n=k}^{\infty} \frac{(2km+1)^{3n+1}}{3n+1} \frac{[N^{-2/3}(2km-1)]^n}{1 - N^{-2/3}(2m-1)} \right\} \\ &\leq 1 + \lambda_H^m \left\{ N^2 + N^{2k/3} \sum_{n=k}^{\infty} \frac{2km+1}{(3n+1)[1 - N^{-2/3}(2km-1)]} [N^{-2/3}(2km+1)^4]^n \right\}. \end{aligned} \quad (65)$$

We then pick $m = (1/4k)N^{1/6}$, so that $N^{-2/3}(2km+1)^4 \leq 1/2$ and

$$\begin{aligned} |\lambda| &\leq (E_{1,k} - 1)^{1/m} \leq N^{2/m} \lambda_H (1 + \mathcal{O}(1))^{1/m} \\ &= \lambda_H (1 + \mathcal{O}(\log(N)kN^{-1/6})). \end{aligned} \quad (66)$$

Using Markov's inequality, the probability that $|\lambda|$ is greater than $c(1 + \mathcal{O}(k \log(N)N^{-1/6})\lambda_H(D))$, for any $c \geq 1$, is bounded by $c^{-(1/4k)N^{1/6}}$. ■

IV. DISCUSSION

We have introduced quantum and classical tensor product expanders. These provide a way to approximate t -designs by acting many times with a small number of unitaries. An important open question is whether efficient implementations of these tensor product expanders exist.

Acknowledgments AWH thanks Richard Low for catching an error in the proof of Theorem 4, as well as useful discussions about Lemma 1. MBH thanks the KITP for hospitality while some of this research was completed. MBH was supported in part by the National Science Foundation under Grant No. PHY05-51164 and supported by U. S. DOE Contract No. DE-AC52-06NA25396. AWH was supported by the European Commission under a Marie Curie Fellowship (ASTQIT, FP-022194), the integrated EC project “QAP” (contract no. IST-2005-15848), the U.K. EPSRC, project “QIP IRC” and the Army Research Office under grant W9111NF-05-1-0294.

-
- [1] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and Approximate Unitary 2-Designs: Constructions and Applications. arXiv:quant-ph/0606161, 2006; D. Gross, K. Audenaert, and J. Eisert. Evenly distributed unitaries: on the structure of unitary designs *J. Math. Phys.* **48**, 052104, 2007. arXiv:quant-ph/0611002.
 - [2] R. F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model *Phys. Rev. A* **40**, 4277, 1989; C. H. Bennett et. al. Purification of Noisy Entanglement and Faithful Teleportation via Noisy Channels. *Phys. Rev. Lett.* **76**, 722, 1996. arXiv:quant-ph/9511027; C. H. Bennett, D. P. Divincenzo, J. A. Smolin, W.K. Wootters. Mixed State Entanglement and Quantum Error Correction. *Phys. Rev. A* **54**, 3824, 1996. arXiv:quant-ph/9604024.
 - [3] M. B. Hastings. Entropy and Entanglement in Quantum Ground States. *Phys. Rev. B* **76**, 035114, 2007. arXiv:cond-mat/0701055.
 - [4] A. Ben-Aroya and A. Ta-Shma. Quantum expanders and the quantum entropy difference problem. arXiv:quant-ph/0702129, 2007.
 - [5] A. Ben-Aroya, O. Schwartz and A. Ta-Shma. An explicit construction of quantum expanders. arXiv:quant-ph/0709.0911; A. Ben-Aroya, O. Schwartz and A. Ta-Shma. Quantum Expanders: Motivation and Constructions. *Proc. of 2008 IEEE CCC*, pp. 292–303, 2008.
 - [6] J. Eisert and D. Gross. Quantum Margulis expanders. *Q. Inf. Comp.*, vol. 8, pp. 722–733, 2008. arXiv:0710.0651.
 - [7] A. W. Harrow. Quantum expanders from any classical Cayley graph expander. *Q. Inf. Comp.*, vol. 8, no. 8/9, pp. 715–721, 2008. arXiv:quant-ph/0709.1142.
 - [8] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.* **43**, 439, 2006.
 - [9] J. Friedman. A proof of Alon’s second eigenvalue conjecture and related problems. *Proc. of 35th annual STOC*, pp. 720–724, 2003. arXiv:cs/0405020.
 - [10] G. Aubrun. A remark on the paper ‘Randomizing quantum states: constructions and applications’. arXiv:0802.4193, 2008.
 - [11] P. Hayden, D. W. Leung, P. W. Shor and A. J. Winter. Randomizing quantum states: constructions and applications. *Commun. Math. Phys.* vol. 250, no. 2, pp. 371–391, 2004. arXiv:quant-ph/0307104.
 - [12] P. Dickinson and A. Nayak. Approximate Randomization of Quantum States With Fewer Bits of Key. *AIP Conference Proceedings*, **864**, 18, 2006. arXiv:quant-ph/0611033
 - [13] A. Ambainis and A. Smith. Small Pseudo-Random Families of Matrices: Derandomizing Approximate Quantum Encryption. *Proc. RANDOM 2004*, pp. 249–260, 2004. arXiv:quant-ph/0404075.
 - [14] A. Lubotsky, R. Phillips and P. Sarnak. Hecke operators and distributing points on the sphere I. *Comm. Pure Appl. Math.* **39**, Supplement I, S149–S186, 1986; A. Lubotsky, R. Phillips and P. Sarnak. Hecke operators and distributing points on the sphere II. *Comm. Pure Appl. Math.* **40**, 401–420, 1987.
 - [15] C. Zalka. Implementing high dimensional unitary representations of $SU(2)$ on a Quantum Computer. arXiv:quant-ph/0407140, 2004.
 - [16] A. Brodsky and S. Hoory. Simple Permutations Mix Even Better. arXiv:math/0411098; S. Hoory, A. Magen, S. Myers, and C. Rackoff. Simple Permutations Mix Well. *Proc. 31st ICALP*, 2004.
 - [17] M. Kassabov. Symmetric groups and expanders. *Inventiones Mathematicae*, vol. 170, n. 2, 2007. arXiv:math.GR/0505624
 - [18] A. Brodir and E. Shamir. On the second eigenvalue of random regular graphs. *Proc. 28th FOCS*, pp. 286–294, 1987.
 - [19] M. B. Hastings. Random unitaries give quantum expanders. *Phys. Rev. A* **76**, 032315, 2007. arXiv:0706.0556
 - [20] G. Tóth and J. J. García-Ripoll. Efficient algorithm for multi-qudit twirling for ensemble quantum computation. *Phys. Rev. A* **75**, 042311, 2007. arXiv:quant-ph/0609052.
 - [21] A. Y. Kitaev, A.H. Shen and M.N. Vyalii. *Classical and Quantum Computation*. American Mathematical Society, Providence, RI. 2002.
 - [22] A. W. Harrow, B. Recht and I. L. Chuang. Efficient Discrete Approximations of Quantum Gates. *J. Math. Phys.* **43**, 4445 (2002). arXiv:quant-ph/0111031
 - [23] J. Bourgain and A. Gamburd. New results on expanders. *C. R. Acad. Sci. Paris, Ser. I.* vol. 342, pp. 717–721, 2006.
 - [24] R. Goodman and N. R. Wallach. *Representations and invariants of the classical groups*, 1998.
 - [25] A. Barenco, A. Berthiaume, D. Deutsch, A. Ekert, R. Jozsa and C. Macchiavello. Stabilization of quantum computation by symmetrization. *SIAM J. Comput.* vol. 26, no. 5, pp. 1541–1557, 1997. arXiv:quant-ph/9604028

[26] A. Gamburd, D. Jakobson and P. Sarnak. Spectra of elements in the group ring of $SU(2)$. J. Eur. Math. Soc., vol. 1, pp. 51–85, 1999.

APPENDIX A: PROOF OF LEMMA 1

First, we reduce to the case when the matrices are 2×2 with $\Pi = |1\rangle\langle 1|$ and X is diagonal. Express $\|pX + (1-p)Y\|$ as the maximum of $\langle \psi | pX + (1-p)Y | \psi \rangle$ over all unit vectors $|\psi\rangle$. Write $|\psi\rangle$ as $|\psi\rangle = \cos(\theta)|\psi_1\rangle + \sin(\theta)|\psi_2\rangle$, where $0 \leq \theta \leq \pi/2$ and $|\psi_1\rangle, |\psi_2\rangle$ are normalized vectors such that $\Pi|\psi_1\rangle = |\psi_1\rangle$ and $(I - \Pi)|\psi_2\rangle = |\psi_2\rangle$. Our conditions on X imply that $\langle \psi | X | \psi \rangle = \cos^2(\theta) + \langle \psi_2 | X | \psi_2 \rangle \sin^2(\theta)$ and that $|\langle \psi_2 | X | \psi_2 \rangle| \leq 1 - \epsilon_X$. Next, for $i, j = 1, 2$ define $Y_{i,j} = \langle \psi_i | Y | \psi_j \rangle$. Since $\|Y\| \leq 1$, we also have that $\|\sum_{i,j=1}^2 Y_{i,j}|i\rangle\langle j|\| \leq 1$. We can now replace Y with $\sum_{i,j=1}^2 Y_{i,j}|i\rangle\langle j|$ and X with $|1\rangle\langle 1| + \langle \psi_2 | X | \psi_2 \rangle |2\rangle\langle 2|$.

Now suppose that $|\langle \psi | X | \psi \rangle| \geq 1 - \epsilon_X \epsilon_Y / 12$. Using our bound on $|\langle \psi_2 | X | \psi_2 \rangle|$, we obtain

$$1 - \frac{\epsilon_X \epsilon_Y}{12} \leq \cos^2(\theta) + \sin^2(\theta)(1 - \epsilon_X) = 1 - \sin^2(\theta)\epsilon_X,$$

implying that $\sin^2(\theta) \leq \epsilon_Y / 12$. We will show that this yields an upper bound on $\langle \psi | Y | \psi \rangle$.

Since $\|Y\| \leq 1$, we have $|Y_{1,2}|, |Y_{2,1}| \leq \sqrt{1 - |Y_{1,1}|^2}$. Thus

$$|\langle \psi | Y | \psi \rangle| \leq \cos^2(\theta)|Y_{1,1}| + \sin(\theta)\cos(\theta)(|Y_{1,2}| + |Y_{2,1}|) + \sin^2(\theta)|Y_{2,2}| \quad (\text{A1})$$

$$\leq \cos(\theta)|Y_{1,1}| + \sin(\theta)2\sqrt{1 - |Y_{1,1}|^2} + \frac{\epsilon_Y}{12}. \quad (\text{A2})$$

If θ were not constrained then the first two terms of (A2) would be maximized by taking θ to be $\hat{\theta} = \arctan(2\sqrt{1 - |Y_{1,1}|^2}/|Y_{1,1}|) \geq \arctan(2\sqrt{2\epsilon_Y - \epsilon_Y^2}/(1 - \epsilon_Y)) \geq \arctan(2\sqrt{2\epsilon_Y})$. Using $\sin^2(\arctan(z)) = z^2/(1 + z^2)$, we have $\sin^2(\hat{\theta}) \geq 8\epsilon_Y/(1 + 8\epsilon_Y) \geq \epsilon_Y/2$. Since θ is constrained to lie in $[0, \arcsin(\sqrt{\epsilon_Y/12})]$, it cannot equal $\hat{\theta}$. Thus maximizing (A2) will require setting θ to one of the endpoints of the allowed region. In particular, the maximum value of (A2) occurs when $\sin^2(\theta) = \epsilon_Y/12$. A similar argument proves that setting $|Y_{1,1}| = 1 - \epsilon_Y$ maximizes (A2) as well. Now we calculate

$$|\langle \psi | Y | \psi \rangle| \leq (1 - \epsilon_Y) + 2\sqrt{\frac{\epsilon_Y}{12}}\sqrt{2\epsilon_Y - \epsilon_Y^2} + \frac{\epsilon_Y}{12} \leq 1 - \left(1 - \sqrt{\frac{2}{3}} - \frac{1}{12}\right)\epsilon_Y \leq 1 - \frac{\epsilon_Y}{10} \quad (\text{A3})$$

We have shown that for any ψ , either $\langle \psi | X | \psi \rangle \leq 1 - \epsilon_X \epsilon_Y / 12$ or $\langle \psi | Y | \psi \rangle \leq 1 - \epsilon_Y / 10$. We now use the triangle inequality to bound

$$\begin{aligned} \langle \psi | pX + (1-p)Y | \psi \rangle &\leq \max\left(p\left(1 - \frac{\epsilon_X \epsilon_Y}{12}\right) + (1-p), p + (1-p)\left(1 - \frac{\epsilon_Y}{10}\right)\right) \\ &\leq 1 - \frac{\epsilon_Y}{12} \min(p\epsilon_X, 1-p). \end{aligned}$$

Since this bound applies for all normalized $|\psi\rangle$, it must also upper-bound $\|pX + (1-p)Y\|$. Thus we obtain (30). The remaining steps of the Lemma are direct calculations. \blacksquare